DOCUMENT RESUME

ED 045 715 TM 000 286

AUTHOR	Smith, Richard; Fiedler, Fred E.
TITLE	The Measurement of Scholarly Work in Academic Institutions: A Critical Review of the Literature.
INSTITUTION	Washington Univ., Seattle. Dept. of Psychology.
SPONS AGENCY	Office of Naval Research, Washington, D.C.
REPORT NO	TR-70-2
PUB DATE	Feb 70
NOTE	16p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$0.90
DESCRIPTORS	*Academic Rank (Professional), Academic Standards, Citation Indexes, College Faculty, Colleges, *Departments, *Faculty Evaluation, *Performance Criteria, *Professional Recognition, Publications, Research, Universities

ABSTRACT

This critical review of the literature is concerned with the measurement of scholarly work done by the faculties of universities and colleges. Such measures of output as individual and departmental ratings by scholars, the amount of recognition awarded, the number of publications written, and the number of citations to published work, are discussed and compared. Reference is made to studies that present empirical findings relating these measures to one another. It is concluded that among the alternatives discussed, the citation count is the least biased measure of scholarly work in academic institutions. (Author)

EDO 45715 AD 704516

ORGANIZATIONAL RESEARCH

DEPARTMENT OF PSYCHOLOGY • UNIVERSITY OF WASHINGTON, SEATTLE, WASHINGTON

THE MEASUREMENT OF SCHOLARLY WORK IN ACADEMIC INSTITUTIONS:

A CRITICAL REVIEW OF THE LITERATURE

Richard Smith and Fred E. Fiedler

University of Washington

Seattle, Washington

Technical Report No. 70-2

February 1970

TM 000 286

This document has been approved for public release and sale; its distribution is unlimited.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED 045715

ORGANIZATIONAL RESEARCH
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON

THE MEASUREMENT OF SCHOLARLY WORK IN ACADEMIC INSTITUTIONS:

A CRITICAL REVIEW OF THE LITERATURE

Richard Smith and Fred E. Fiedler

University of Washington
Seattle, Washington

Technical Report No. 70-2

February 1970

Contract N0014-67-A-0103-0012
Office of Naval Research
Department of the Navy

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

THE MEASUREMENT OF SCHOLARLY WORK IN ACADEMIC INSTITUTIONS: A CRITICAL REVIEW OF THE LITERATURE

Richard Smith and Fred E. Fiedler
University of Washington, Seattle, Washington

The past three decades have seen a phenomenal expansion of higher education in the United States. While only six to twelve percent of the college-aged youths in Europe are in institutions of higher learning, in the United States approximately fifty percent of the college-aged population are attending colleges and universities. It is very likely that we will see a further increase in the proportion of young men and women in college within the next few years. This sudden expansion of the college and university populations, in conjunction with the current tightening of state and federal funds for higher education, will undoubtedly result in closer scrutiny of the effectiveness with which educational institutions meet society's goals as well as their own objectives. It is highly appropriate, therefore, that social scientists and educators concern themselves with the adequacy of our educational systems. The cornerstone of systematic research in this field that can lead to meaningful educational reforms must be an adequate basis for evaluating performance. The present paper reviews some measures of organizational performance in higher educational institutions.

We shall be particularly concerned with criteria of scholarly performance in graduate teaching and research: i.e., the generation and dissemination of knowledge. These are clearly two major purposes of colleges and universities. While these institutions have such other important functions as employing academic and nonacademic personnel, socializing young adults, and providing highly specialized manpower for government and industry, this paper confines itself to the measurement of the academic excellence of university departments and individual faculty members.

Reputational Measures of Departmental Eminence. The earliest major attempts to measure the performance of academic departments were those undertaken by R. M. Hughes (Robertson, 1928; Hughes, 1934). College and

university professors were asked to rate the quality of graduate departments; 20 fields were rated in 1925 and 35 fields in 1934. Keniston (1959) obtained similar rankings by asking department chairmen in 28 fields to rate sister departments in other universities. This was followed by the American College of Education (ACE) study in 1964 (Cartter, 1966), which asked a wide range of scholars to rate the quality of 106 university faculties in 29 different fields. The raters also judged the attractiveness of the graduate programs provided by these different departments. The resulting rankings of departments, all providing training at the doctoral level, comprise at this time the best known index of departmental performance. (A revised ranking of departments is expected to appear within the near future.)

Although measures of this type have some obvious advantages and also, as we shall show later, moderate validity, they also have some obvious shortcomings. The major limitation is the high degree of halo effect from which the department benefits (or suffers) as a result of being part of a well- or poorly-known university. In general, good departments tend to be located in good universities; however, some excellent departments can be found in less highly regarded universities, and some departments at outstanding universities may be quite poor. And even when the halo effect is not present, it is possible for a rater's judgment to be influenced by misinformation, hearsay, and his own personal biases.

A second major limitation of reputational measures is the considerable time-lag between actual changes in a department's personnel and teaching program, and the reflection of these changes in ratings by scholars at other schools. Eminent scholars are notoriously mobile, and it is by no means rare that a department is suddenly stripped of the four or five outstanding scholars on which its reputation has been built.

Finally, reputational measures appear to be unduly influenced by the size of the department: a large department is likely to be more visible than a small department. Fiedler and Biglan (1969), in a study of academic departments of the University of Illinois, found a correlation of .54 between ACE rating and number of faculty members in the department. It may be argued, therefore, that reputational measures are based, at least in part, on departmental visibility. When the visibility is based upon the excellence of

the research by members of the faculty and the outstanding students they have produced, it quite appropriately contributes to departmental reputation. But if the visibility is based upon the visibility of the university or the sheer size of the department, the reputational measures will produce spurious results.

Measures Reflecting Individual Faculty Member Performance. A number of studies have been published which define academic performance in terms of faculty productivity at the individual level. Since departmental measures are obtained essentially by summing or averaging individual measures, and since even the ACE ratings are in effect based on an averaging process in the rater's head, using an unspecified weighting system, it is quite important to develop individual measures of performance. Furthermore, ratings of individual faculty members have the advantage of making explicit the contributions made by various members of the department.

Most ratings of individual faculty members are based on their publications. It must be borne in mind, however, that the basis for making these ratings is less direct than would appear at first glance. It is rare indeed that the rater is fully acquainted with an individual's writing, and even more unusual for the rater to have read all or even most of the rater's publications. Thus, we are dealing again with a measure of reputation.

Since it is reasonable to assume that researchers (as practically all other people) strive to be rewarded for their work, one way of measuring research performance is to consider the distribution of rewards by the academic community. The main reward is recognition (Merton, 1957), a term which encompasses rewards of varying importance. The highest form of recognition a man can receive is to have something named after him: Euclidean geometry, Newtonian mechanics, Lewinian theory, the Wigner effect. Only a small number of scholars is ever recognized in this fashion. Considerably greater numbers are awarded prizes and awards for their work, and some are granted membership in select societies. There are various recognitions of eminence including consultantships, selection to editorial boards, scientific panels, or advisory boards, as well as election to office in scientific or scholarly societies.

Crane (1965) related research productivity and recognition in the departments of biology, political science, and psychology of three universities--one prestigious, one intermediate, and one low in prestige.

Recognition was measured by such honors as the presidency of, or membership in, certain associations and societies, honorary degrees, postdoctoral fellowships, service on journal editorial boards, and other prizes. The measure of productivity was the number of publications, with books given the weight of four journal articles. The study showed that a man's recognition was highly related to the prestige of his current academic affiliation. Of somewhat less importance was the eminence of the man's former academic sponsor. Continuity of research was also related to recognition, provided the work was conducted at a major university. Crane also found that 56 percent of the highly productive scientists she studied had won recognition, whereas only 30 percent of the less productive had been so recognized. She concludes that affiliation with a major university is more likely to lead to recognition for a scientist than is high productivity or sponsor prestige.

Crane's study casts doubt on the adequacy of recognition as a performance measure. The prestige of a man's university or department apparently facilitates recognition of a man's research. Moreover, recognition measures are of limited usefulness since there are many scholars who receive little or no public recognition of the type incorporated in Crane's index--and probably most indices of recognition that can be developed. It is also likely that in some instances, personal biases unduly influence the awarding of recognition in the academic community.

Quantity of Research Publications. Lipetz (1965) has argued that scientific achievement can best be assessed by measuring the scientific content of research, as presented in the scientist's written communications. In effect, Lipetz calls for a content analysis of journal articles, books, and technical reports. A simpler measure of an individual's scholarly output is the number of articles, books, and reports he has published. A numerical count of publications is the most widely used and notorious method for quickly assessing an academician's productivity. Thus, Somit and Tanenhaus (1964) assert that the quantity of publication is the "standard by which merit is measured" in political science. Harmon (1963) found correlations of .61 and .76 between publication and a rating criterion of individual physical scientists and biological scientists, respectively. Meltzer (1949) found that the number of publications correlated .20 with the eminence of the institution granting the PhD, and that a poorly conceived measure of individual repute correlated .27 with eminence. (The measure of repute is poorly conceived

because the sample of 266 was divided into only two categories, high and low, and three-fourths of the 266 were placed in the low category.) Manis (1951) found a correlation of .28 between the same measure of repute and the eminence of the individual's current department as indicated by Hughes' 1934 study, and a correlation of .18 between eminence and the sheer quantity of an individual's publications. Clark (1957) reported a correlation of .47 between Psychological Abstract items (a measure of quantity) and a rating of individual "eminence." On the other hand, Fiedler and Biglan (1969) found a slightly negative correlation between American Council of Education ratings of a department's quality and the average number of books published by members of the department (-.18, N=25). In contrast, the correlation between the average number of journal articles published in a department and the department's ACE rating was slightly positive (.38, N=25). Cartter (1966) found strong relationships between amount of publication and ACE ratings of political science and economics departments, but a somewhat lower relationship for English departments. He did not report relationships for physical science departments, for which Fiedler and Biglan reported correlations around zero. The latter finding suggests that the relations between reputation and departmental productivity (and probably individual productivity) as measured by number of publications may vary widely from discipline to discipline, or among families of disciplines.

A quantity measure of performance has its own limitations. The most obvious of these is that a poorly conceived paper published in a badly-edited journal will count as much as will a major contribution to the field which is published in a well-refereed journal. (Indeed, some scholars may produce several mediocre publications per year, thus acquiring a very high publication count.) Second, it is difficult to assign an a priori weighting system. Crane counted a book as equivalent to four journal articles. Meltzer, claiming that an article is equivalent to a chapter, and that there are on the average, 18 chapters per book, used a ratio of 18 to 1.

A good criterion of academic performance obviously should reflect quality as well as quantity. This is not to say that quantity is unimportant. A scholar who rarely publishes will not have the impact someone will have who publishes the equivalent ideas in several different journals and

other publication outlets. Moreover, publication norms differ widely from field to field. While articles in many chemical journals are quite short and some eminent scholars can claim authorship of several hundred articles, papers are more difficult to write in such fields as philosophy or theoretical physics.

Measures of Quality. Cole and Cole (1967) used a criterion of research output that seems to reflect quality more than the publication count does. Unlike Crane, they considered recognition a reward for quality rather than a direct measure of it. Their criterion measure is the number of citations an individual's work receives in the literature during a given number of years. Although they were not the first to describe such a measure (see Ruja, 1956; Clark, 1957; Myers and DeLevie, 1966; see also Ornstein, et al., forthcoming*), their study is certainly among the most significant on the subject of research output. Cole and Cole studied 120 physicists in the United States, using the average number of weighted citations to a physicist's research in his three most heavily cited years. A citation was given more weight if it was a reference to an older piece of research, since most citations are to recent work. According to this rationale, a scholar deserves extra credit if his 15-year-old research is still worth quoting.

This citation measure has several advantages. It is not greatly influenced by quantity, since a few published papers by a man might be so outstanding that they become a benchmark for later research (e.g., Einstein's small monograph on his special theory of relativity; Darwin's Origin of Species). Quantity of publication can be systematically eliminated from the measure by dividing the number of citations by the number of publications over a certain period of time. An index of citations is relatively easy to obtain for certain fields for which the Science Citation Index is available, though the routine labor required for publication indices for all academic fields might require prodigious work. The index is based upon evaluations of research rather than on evaluations of persons; and finally, a large number of a man's colleagues have a choice of citing or not citing his work, and hence a voice in the outcome. In a sense, a citation is a rating; a citation implies that the writer considers the cited work significant enough so that it has to be taken into consideration. A citation is therefore an "unobtrusive measure" (Webb, et al., 1966) reflecting the impact or significance of a man's work. This is the case even when the reference is critiqued.

*This work has been cited by Cole & Cole, 1967.

The measure does have flaws, however. A significant piece of research may not be recognized for a considerable period of time (consider, for example, Mendel's classic paper on the genetics of the sweetpea). At the other extreme, a piece of research may become so famous that it enters the public domain and is no longer cited by name (e.g., Student's t). Moreover, the differences in fields must be taken into consideration. A man publishing in the area of analytic chemistry faces different competition than does a man in the area of Hittite mythology or Urdu grammar. Finally, a researcher frequently has a choice of sources he might cite to support his propositions. In these cases he is more likely to refer to an eminent and widely known authority working in a major university than to a relatively unknown researcher at a small and undistinguished college, even though the latter might provide somewhat stronger support for his case. The prestige of a man or of his university is also likely to influence an editor's decision whether or not to accept a paper for publication. Almost anything a Nobel laureate might wish to write is likely to be published by a professional journal, even though the paper may not be up to the journal's usual standards.

The above criticisms notwithstanding, a measure based on citations may provide as unbiased a measure of the quality of a man's work as we are likely to get. Cole and Cole provide some supporting data. They designated the publication of 30 papers over a three-year period as the cutting score between high and low quantity of output; they considered 60 citations as the cutting score between high and low quality. They then classified physicists into four categories: the prolific (high quantity-high quality); the mass producers (high quantity-low quality); the perfectionists (low quantity-high quality); and the silent physicists (low quantity-low quality). Quality and quantity measures were also correlated with various indices of recognition. (See table, page 8.)

The Coles' data show that the quality index correlates more highly with measures of recognition than does the quantity index. The correlation between the ACE rank of a man's department and the number of his awards was .50, while the quality and quantity of the individual's output correlated .72. According to Cole and Cole, quality, not quantity, is the main factor distinguishing award-winning physicists from those who have not

COEFFICIENTS OF CORRELATION BETWEEN QUANTITY AND QUALITY

OF RESEARCH AND THREE MEASURES OF RECOGNITION*

Quality and Quantity of Research	Measures of Recognition: Awards: Prestige of highest award	Measures of Recognition: Awards: Number of awards	Measures of Recognition: Rank of Department	Measures of Recognition: Percent of Community of Physicists Familiar with Individuals' Research
1. Quantity	.35	.46	.24	.49
2. Number of papers per year	.28	.32	.19	.43
3. Quality	.41	.67	.33	.64

*Reprinted from Cole & Cole (1967).

been so rewarded. Quality was found to account for 44 per cent of the variance in the number of awards, and adding the factor of quantity did not increase the amount of variance accounted for.

Quality was also more important than quantity in distinguishing physicists in the top ten departments from those elsewhere. In fact, there were more perfectionists than prolific physicists at the top ten departments (based on ACE ranking). These data do not seem to support the publish-or-perish doctrine when it is interpreted primarily in terms of number of publications.

As one might expect, physicists high on the quality index are better known by their colleagues than are physicists low on the index. The number of citations to a physicist's work is related to the number of other physicists who have read at least some of his work and also to the number who have heard of him but have not read his work.

Bayer and Folger (1966) also utilized a measure based on the number of citations as an index of quality. They studied 467 biochemists who had received their degrees in 1957 or 1958 and found a significant correlation of .21 between quality of graduate school (based on ACE ratings) and the number of citations. The correlation between IQ and number of citations was -.05, however. (Obviously this correlation is highly attenuated by restriction of range.) The data suggest that the quality of graduate education may be important in determining future research performance, although the school's selection of students as well as the self-selection of applicants for a particular school make such an interpretation very tenuous.

Various other studies tend to support the Cole and Cole findings. Clark found a correlation of .47 between quantity and ratings of individual eminence for psychologists and a correlation of .68 between citation count and eminence. He found lower correlations between eminence and number of offices held in the American Psychological Association or number of Ph.D. students.

Pelz and Andrews (1966) conducted a major study on the productivity of scientists. Among their findings is one which is particularly relevant to the present discussion: A correlation of .39 between the number of papers published by scientists in research laboratories and their "scientific contribution" as rated by their supervisors and colleagues.

As we have pointed out above, Cole and Cole found that departmental prestige and the eminence of the individual faculty member are closely interrelated. Which is cause, and which is effect is, of course, difficult to fathom.

Prestigious departments attract eminent scholars, and eminent scholars contribute to the distinction of their department. In addition, the climate of an eminent department undoubtedly contributes to the quality of research conducted in the department. Wilson (1943) studied the prestige patterns in the academic community by asking a large sample of scholars in each of 12 fields to name the 20 most important contributors to their field. He found that 90.6 percent of the 120 leading men (ten from each field) were at ten highly prestigious universities. This finding supports Cole and Cole's correlation of .50 between the number of awards and departmental prestige. If we consider prestige a correlate of the quality of a man's scholarly research, it is clear that the fate of the individual and that of his academic community are very closely interwoven.

Discussion. The data presented in this paper fairly well speak for themselves. We have attempted to review and bring together findings from key studies which examine the correlates of scholarly research output. When the data are considered as a whole, it would appear that quantity of publication is moderately related to individual or departmental eminence· that productivity and recognition are moderately related; that citation counts correlate well with recognition and with individual eminence, less well with departmental prestige. The relationship between citation counts and quantity of publication is less clear: Cole and Cole report a correlation of .72, but Clark offers .47 for total Psychological Abstract counts correlated with citations and only .36 for a four-year period of abstract counts.

Of the indices that are currently available, the measures based on citation seem to be least contaminated by such factors as the prestige of the man's department or university, or the sheer number of papers he has published. While measures reflecting the number of citations have their own problems, not the least of which is the amount of work which they require, it should be possible to reduce considerably some of the required effort. Many sciences are represented in the Science Citation Index. For other fields, one might take citations in standard texts, handbooks, annual reviews, and journals critically reviewing the literature as an acceptable approximation. It should also be possible to develop intermediate measures of output. Research on this problem is currently in progress by the writers and their colleagues.

References

Bayer, A. E., & Folger, J. Some correlates of a citation measure of productivity in science. Sociology of Education, 1966, 39, 381-390.

Cartter, A. M. An Assessment of Quality in Graduate Education, American Council on Education, Washington, D. C., 1966.

Clark, K. E. America's Psychologists, American Psychological Association, Washington, D. C., 1957.

Cole, S., & Cole, J. R. Scientific output and recognition: A study in the operation of the reward system in science. American Sociological Review, 1967, 32, 377-390.

Crane, D. Scientists at major and minor universities: A study of productivity and recognition. American Sociological Review, 1965, 30, 699-714.

Fiedler, F. E., & Biglan, A. The First Job after Graduation as a Measure of Departmental Performance. Technical Report 74, April 1969, Group Effectiveness Research Laboratory, Urbana, Ill.

Harmon, L. R. The development of a criterion of scientific competence. In Taylor & Barron (Ed.), Scientific Creativity, Wiley, New York, 1963.

Hughes, R. M. Report of the Committee on Graduate Instruction. Educational Record, April 1934, 192-234.

Keniston, H. Graduate Study and Research in the Arts and Sciences at the University of Pennsylvania. University of Pennsylvania Press, Philadelphia, 1959.

Lipetz, B. The Measurement of the Efficiency of Scientific Research, Intermedia, Inc., Carlisle, Mass., 1965.

Manis, J. G. Some academic influences upon publication productivity. Social Forces, 1951, 29, 267-272.

Meltzer, B. N. The productivity of social scientists. American Journal of Sociology, 1949, 55, 25-29.

Merton, R. K. Priorities in scientific discovery: A chapter in the sociology of science. American Sociological Review, 1957, 22, 635-659.

Myers, R. A., & DeLevie, A. S. Frequency of citation as a criterion of eminence. Journal of Counseling Psychology, 1966, 13, 245-246.

Ornstein, L., et al. Research, Scholarly Publication and Teaching: The Development of Some Objective Measures of a Scientist's Impact on Science. Mount Sinai School of Medicine, forthcoming.

Pelz, D. C., & Andrews, F. M. Scientists in Organizations. Wiley, New York, 1966.

Robertson, D. A. (Ed.) American Universities and Colleges. Scribner, New York, 1928, pp. 161-163.

Ruja, H. Productive psychologists. American Psychologist, 1956, 11, 148-149.

Somit, A., & Tanenhaus, J. American Political Science. Atherton Press, New York, 1964.

Webb, E. J., et al. Unobtrusive Measures. Rand McNally, Chicago, 1966.

Wilson, L. Prestige patterns in scholarship and science. Southwestern Social Science Quarterly, 1943, 23, 305-319.

UNCLASSIFIED

DD FORM 1473

DOCUMENT CONTROL DATA - R&D

1. ORIGINATING ACTIVITY

Organizational Research
Department of Psychology
University of Washington, Seattle, Washington

2. REPORT SECURITY CLASSIFICATION

Unclassified

3. REPORT TITLE

The Measurement of Scholarly Work in Academic Institutions: A Critical Review of the Literature

4. DESCRIPTIVE NOTES

Technical Report 70-2, February 1970

5. AUTHORS

Richard Smith and Fred E. Fiedler

6. REPORT DATE

February 1970

7a. TOTAL NO. OF PAGES

14

7b. NO. OF REFERENCES

21

8a. CONTRACT OR GRANT NO.

N00014-67-A-0103-0012

8b. PROJECT NO.

177-472

9a. ORIGINATOR'S REPORT NUMBER

TR 70-2

10. AVAILABILITY/LIMITATION NOTICES

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

11. SUPPLEMENTARY NOTES: NONE

12. SPONSORING MILITARY ACTIVITY

Office of Naval Research

13. ABSTRACT

This critical review of the literature is concerned with the measurement of scholarly work done by the faculties of universities and colleges. Such measures of output as individual and departmental ratings by scholars, the amount of recognition awarded, the number of publications written, and the number of citations to published work, are discussed and compared. Reference is made to studies that present empirical findings relating these measures to one another. The authors conclude that, among the alternatives discussed, the citation count is the least biased measure of scholarly work in academic institutions.

14. KEY WORDS

Academic performance
Research Output
Scholarly performance
Academic criteria
Journal citations
Academic eminence
Journal publications
Publications, quality
Publications, quantity